DOCUMENT RESUME

ED 240 901 HE 017 000

AUTHOR Prather, James E.; And Others
TITLE Evaluations by Students in Mathematics Courses of the Effectiveness of Teaching.
PUB DATE 25 Jan 84
NOTE 13p.; Paper presented at the Annual Meeting of the American Mathematical Society (Louisville, KY, January 25, 1984).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *College Mathematics; Higher Education; *Mathematics Instruction; *Predictor Variables; Reliability; *Student Evaluation of Teacher Performance; *Teacher Effectiveness; Validity

ABSTRACT

The extent to which students' ratings of teacher performance in mathematics courses are reliable, valid, and unbiased was investigated. Data on 746 courses by 52 faculty members over a 4-year period were used in a multiple regression analysis. It was found that students gave higher evaluations to faculty members when the students had previously taken a course with the faculty member. This evidence suggests that the measure of students' "repeating" with faculty members can be a valid indicator of teaching effectiveness. However, student ratings and the repeat measure are subject to the same biases. Teacher personality, student motivation, and orientation could affect the repeat measure in a manner similar to their effect on student ratings of teacher effectiveness. The variables having a moderately strong, positive association with students' ratings include: average number of repeats, percent of lecture courses taught, average grades in the course, number of previous courses taken, percent of lower level courses taught, total average grade point average (GPA) and percent of reading and research tutorials taught. Negative factors are: average GPA of students in the course, percent of students withdrawing, attendance rate of students, number of courses taught, class size, and percent of courses taught at upper level. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED240901

Evaluations by Students in Mathematics Courses of The Effectiveness of Teaching

James E. Prather

Fred A. Massey

James E. Greene, Jr.

Joseph S. Sturgeon

Georgia State University
University Plaza
Atlanta, Georgia

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

James E. Prather

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

HE 017 000

Paper Presented at
American Mathematical Society
Louisville, Kentucky
January 25, 1984

Evaluations by Students in Mathematics Courses of the Effectiveness of Teaching

Many faculty development programs use students' evaluations of instructor performance as a key component. The issue of using student evaluations for faculty personnel decisions has raised great levels of anxiety and vigorous debate. These policy questions cannot be resolved until a broad concensus is achieved about measurement issues that remain outstanding in the area of student evaluations of teaching.

The most frequently addressed measurement problem concerns the question of bias in students' ratings of teacher performance. The removal or control for bias has been regularly discussed in the published literature, though a generally accepted model for minimizing bias in students' evaluations of teaching has not yet been achieved. Less research has been done on the important issue of validity vs. reliability.

Research on evaluations of classroom teaching introduced various measures of teaching effectiveness. While the teachers' self evaluation and peer evaluation have been used with ambigous results, the primary way to measure effectiveness has been with anonymous ratings by students. This system of students' secret scoring of their instructors has itself generated much controversy, especially in the areas of reliability, validity, and bias. The question of reliability of student ratings has been addressed with a certain amount of success (Yunker, 1983). The validity question of students' evaluations of mathematics instructors being possibly biased is addressed in this paper. The bias issue predominates in the literature and apprehensions exist that the measurement process so effects students and teachers that any reading of teacher effectiveness is necessarily distorted. In the research literature it is noted that factors

outside the direct influence of the instructor (whether the course taught is required or is an elective, or how large the class enrollment is) may have a biasing effect on student ratings of teacher performance.

Academic administrators need to accurately gauge teacher performance. This need, combined with problems of reliability, validity, and bias in current measures, underlies the search for alternate ways of measuring teaching performance. In this study, an unobstrusive indicator of teacher performance is employed. This indicator is based on observable student behavior, rather than the anonymous judgments and opinions of students, peers, or administrators. The observable student behavior that is used as an indicator is the number of times students return to a given teacher for additional courses. In large and diverse colleges and universities, students have a great freedom in choosing their particular schedule of courses, and student behavior can exhibit a pattern of avoiding or seeking certain teachers. Those instructors that stimulate interest in mathematics and encourage intellectual development should be indicated by how often they are sought out by the students who are their followers or proteges. Furthermore, this indicator makes use of existing administrative data and has the advantage of being completely unobstrusive.

Literature of Students' Evaluations of Teaching

The literature on students' evaluation of teaching contains numerous studies conducted to determine what factors influence or bias students' ratings. Hoffman & Kremer (1980) investigated the effect of instructor-student congruence of attitudes on instructor ratings. Multiple regression analysis showed that almost 50% of the variance in evaluations could be accounted for by attitudes alone. Instructors' attitudes as perceived by students was the best predictor of evaluations, and a positive course evaluation was associated with students attributing to instructors attitudes preferred by the students themselves.

The personality characteristics of an instructor can have an impact on students' rating. In an analysis of students' written answers to open-ended evaluation questions, Braskamp, Ory, and Pieper (1981) reported that one-fourth of the comments pertained to the instructor's pedagogical skills, while about half of the comments concerned the instructor. Abrami, Perry and Leventhal (1982) researched the relationship between student personality characteristics, teacher ratings, and student achievement. They found no consistent association between ratings and student personality characteristics, but the perceived personality characteristics of faculty were correlated with ratings of teacher effectiveness. Teacher effects on ratings appeared significantly greater than teacher effects on achievement.

There are studies which focus on how student and course variable relate to students' evaluations of teaching. Marsh (1980) examined the relationship between student evaluations and a set of background characteristics. Favorable student ratings were correlated with prior subject interest, higher expected grades, higher levels of workload difficulty, and a higher percentage of students taking the course for general interest only. Overall and Marsh (1980) investigated the relative contribution of course level (undergraduate versus graduate), course type (accounting, economics, etc.), and the specific instructor on students' evaluation. The variance which could be attributed to the specific instructor was much greater than that due to course level or course type. Who teaches a course appeared to be relatively more important than the particular course or the level at which it is taught.

Often researched has been the relationship between grades and instructor ratings. DuCette and Kenny (1982) found that the ratings a student gives to an

instructor and a course are somewhat determined by the grade the student expects to obtain in the course. Abrami, Dickens, Perry, and Leventhal (1980) report that teacher standards for assigning grades had significant effects on student ratings. Howard and Maxwell (1980) argue against the grading leniency bias model--the idea that instructors who are easy graders receive high ratings simply because they are easy graders. They suggest that other variables, especially student motivation, may be more important in determining instructor evaluations. Howard and Maxwell (1982) replicated their earlier study and found no evidence that a grades-influencing-satisfaction interpretation was more likely than its opposite.

A controversy exists in the literature concerning the validity of using students' ratings for faculty personnel decisions. The findings of Marsh (1982) demonstrated agreement between students and instructors on evaluations of teaching effectiveness, and support the validity of student ratings. Dowell and Neal (1982) provide a review of studies which have attempted to link student ratings to student learning as a way of validating student ratings as a measure of teaching ability. The validity of student ratings is quite variable, and is at best only modest. They recommend that student ratings be used with great caution in the processes of faculty review and decision making. Hills (1974) concluded that student ratings of faculty could not be trusted when determining pay increases, promotion, and tenure.

Method

Subjects

Students' evaluations of undergraduate and graduate mathematics courses for a four-year period (1979-1982) are the criteria of interest. A total of 746 courses by 52 different instructors are included in the data base. The courses were for both service and degree major students.

Procedure

Multiple regression was the technique chosen to explore the influence of various factors upon students' tendencies to repeat a given instructor. The level of analysis was the individual course. The independent variables are various academic and demographic characteristics. The "Student Repeats per Course" variable was calculated by tallying the number of times a student had repeated the particular instructor of the course, and dividing this tally by the class size. "Total Average Repeats" was the mean of a given instructor's "Student Repeats per Course" for all courses taught by a given instructor. The measures of student academic performance in mathematics were the average GPA, average grade, percent of students withdrawing from courses, average rate of attendance for the five-year period, and number courses taken previously. Whether the course was graduate or upper level was also included as a variable. Measures for each teacher were included to control for both experience and teaching load of the instructors, these were the number of courses taught and the inverse of the number of students taught. Additional historical measures were characteristics of instructional mode such as the percent of lectures, seminars, theses, and all "other" courses being taught by an instructor. The average class size for each instructor was also included. Instructional level was measured as the percent of graduate, upper, or lower division courses taught. The student demographic characteristics included the percent of minority of males, minority females, nonminority males, and nonminority females.

The dependent variables are measures of teaching performance including if the instructor: 1) was well-prepared; 2) stimulated student thinking; 3) was actively helpful to students; 4) explained course objectives; 5) was fair

and impartial in grading; 6) explained difficult material ; and 7) if the students felt they learned a great deal. An average of the seven items is also included. The multiple regressions for the 746 courses are reported in Table 1 and Table 2. The unstandardized regression coefficients are given in Table 1 and the standardized regression coefficients (Betas) in Table 2.

Results

The R^2's for the 8 measures of students' rating of instructor performance range from .12 to .18 and the adjusted R^2's range from .09 to .15 as shown in Table 1 and table 2 The "Actively Helpful to Students" item had the largest R^2 while "Well-prepared" had the smallest. The indicator of total average number of repeats overall courses taught is among the best predictors of students' ratings. This provides evidence of the validity of the unobtrusive measure of students repeating the same teacher.

The variables having a moderately strong positive association with students' ratings include. average number of repeats, percent of lecture courses taught, overall average grade and average grades in the course, number of courses previous courses taken, percent lower level courses taught, total average GPA, percent "other" course taught (reading and research tutorials) and graduate level course. Negative factors associated with students' ratings are: average GPA of students in the course, percent students withdrawing from the course, attendance rate of students taking the courses, number of courses taught, class size, and percent of course taught at graduate and upper level.

These findings are congruent with previous research that reports positive association of students' rating with higher grades in the course, upper or graduate level course, and smaller class size. The repeat measure as an independent variable is positively associated with the dependent variables at the overall level as well as at the individual course level. The higher Beta's for the overall level of repeats are an indication of the reliability of the repeats measure.

The simple correlations for the rate of repeating same teacher with the eight measures of teaching effectiveness are given in Table 3. These correlations show modest levels of concurrent validity.

Conclusions and Implications

This paper's primary purpose has been to determine the extent to which students' ratings of teacher performance in mathematics courses are reliable, unbiased and valid. Data on 746 courses by 52 faculty over a four-year period were used in a multiple regression analysis.

It was found that students repeating a given faculty member is associated with higher students' evaluations. This evidence suggests that the measure of students' repeating faculty members can be a valid indicator of teaching effectiveness. Student ratings and the repeat measure, however, are subject to the same biases. Teacher personality, student motivaion and orientation could affect the repeat measure in a manner similar to their effect on student ratings of teacher effectiveness.

The ability to use specific administrative data to augment existing systems of evaluating teacher effectiveness is certainly desirable. Diversity of measurement reinforces reliability. Administrative data have the potential to both add much-needed information regarding the evaluation of teacher performance and define additional criteria vital to the assessment of teacher effectiveness.

Table 1

Unstandardized Regression Coefficients for Students' Ratings of Teachers in Mathematics Courses with Selected Instructional, Academic, and Demographic Variables
N=746

	DEPENDENT VARIABLES							
INDEPENDENT VARIABLES	Well-Prepared	Stimulated Student Thinking	Actively Helpful to Students	Course Objectives Explained	Grades Fair/ Impartial	Explains Difficult Material	Students Learn Great Deal	Average of 7 Items
Student Repeats per Course	.08	.11	.12	.06	.05	.14	.11	.10
Student Demographics:								
Percent Minority Female	-.001	.002	-.001	-.001	-.001	-.001		-.001
Percent Minority Male	-.003	-.12	-.004	-.003	-.003	-.005	-.001	-.003
Percent Non-Minority Female	-.001	.003	-.004	-.002	-.002	-.003	.001	-.002
Percent Non-Minority Male	-.002		-.003	-.001	-.002	-.002	-.002	-.002
Student Academic Performance:								
Average Grade	.03	.11	.09	.13	.11	.14	.12	.10
Average GPA	-.09	-.09	-.13	-.20	-.14	-.22	-.05	-.13
Percent Withdrawing	-.001	-.002	-.001	-.002	-.001	-.003	-.001	-.001
Attendance Rate	-.01	-.01	.006	-.008	-.007	-.02	-.02	-.01
No. of Previous Courses Taken	.01	.01	-.008	.01	-.009	.02	.01	.01
Control Variables:								
No. of Courses Taught	-.006	-.004	-.007	-.007	-.002	-.009	-.005	-.006
Inverse No. of Students	-1.26	-.46	-.15	-.64	-.11	-.57	-.41	-.51
Class Size	.001	-.006	-.005	-.004	-.003	-.002	-.006	-.004
Instructional Level:								
Graduate	.07	.20	.13	.12	.05	.15	.09	.12
Upper	.02	.07	-.02	-.04	-.08	.01	.02	
Instructional Data Over 4 Years:								
Instructional Mode:								
Percent Lectures	.05	.05		.04		.07	.07	.08
Percent Seminars	-.18	-.06			.01	-.08		
Percent Thesis			-.02	.02	-.06		.03	.04
Percent Other	.08	.08	.02	.06	.02	.11	.10	.11
Instructional Level:								
Percent Graduate	-.001	-.01				-.01		
Percent Upper		-.006	-.006	-.003	-.004	-.02	-.001	-.002
Percent Lower	.007		.007	.01	.005		.007	.008
Total Average Repeats	.75	.54	.89	.89	.38	1.24	.57	.75
Total Average Grade	.15	.17	.44	.18	.24	.31	.24	.25
Total Average GPA	.37	.18	-.09	.41	.03	.24	.03	.17
Total Average Attendance	-.02	-.005	-.01	-.02	-.009	-.03	.003	-.01
Total Average Previous Courses Taken	.03	-.03	.009	.002	.02	.01	-.02	.004
CONSTANT	-1.97	-1.16	3.65	-1.04	3.68	-3.52	-3.58	-5.29
R^2	.12	.14	.18	.17	.16	.14	.14	.16
Adjusted R^2	.09	.12	.15	.14	.13	.11	.11	.13
S.E.E.	±.45	±.48	±.44	±.40	±.30	±.65	±.47	±.40

Table 2

Standardized Regression Coefficients for Students' Ratings of Teachers in Mathematics Courses with Selected Instructional, Academic, and Demographic Variables
N=746

	DEPENDENT VARIABLES							
INDEPENDENT VARIABLES	Well-Prepared	Stimulated Student Thinking	Actively Helpful to Students	Course Objectives Explained	Grades Fair/ Impartial	Explains Difficult Material	Students Learn Great Deal	Average of 7 Items
Student Repeats per Course	.10	.13	.16	.06	.11	.13	.14	.14
Student Demographics:								
Percent Minority Female	-.03	.04	-.03	-.04	-.04	-.01		-.01
Percent Minority Male	-.05	-.02	-.07	-.06	-.07	-.06	-.01	-.05
Percent Non-Minority Female	-.06	.01	-.14	-.08	-.09	-.09	.007	-.07
Percent Non-Minority Male	-.09		-.11	-.05	-.10	-.07	-.06	-.07
Student Academic Performance:								
Average Grade	.05	.17	.14	.24	.28	.17	.19	.19
Average GPA	-.11	-.09	-.16	-.26	-.25	-.18	-.06	-.17
Percent Withdrawing	-.03	-.005	-.03	-.08	-.02	-.08	-.03	-.05
Attendance Rate	-.16	-.19	-.10	-.12	-.13	-.16	-.19	-.17
No. of Previous Courses Taken	.12	.14	.06	.11	.13	.13	.11	.13
Control Variables:								
No. of Courses Taught	-.22	-.13	-.28	-.31	-.12	-.24	-.19	-.25
Inverse No. of Students	-.12	-.04	-.01	-.07	-.02	-.04	-.04	-.05
Class Size	.02	-.08	-.08	-.07	-.07	-.02	-.09	-.06
Instructional Level:								
Graduate	.05	.13	.10	.09	.05	.07	.06	.09
Upper	.02	.05	-.02	-.04	-.09	.008	.01	
Instructional Data Over 4 Years:								
Instructional Mode:								
Percent Lectures	.56	.49		.45		.54	.69	1.00
Percent Seminars	-.10	-.03			.01	-.03		
Percent Thesis			-.06	.07	-.25		.08	.12
Percent Other	.63	.60	.15	.52	.23	.62	.74	.98
Instructional Level:								
Percent Graduate	-.05	-.30				-.30		
Percent Upper			-.22	-.11	-.17	-.43	-.02	-.08
Percent Lower	.42	-.19	.42	.67	.41		.40	.56
Total Average Repeats	.41	.28	.49	.54	.30	.47	.30	.46
Total Average Grade	.13	.14	.40	.18	.32	.19	.20	.25
Total Average GPA	.25	.11	-.06	.30	.03	.11	.02	.12
Total Average Attendance	-.12	-.02	-.08	-.09	-.07	-.09	.01	-.07
Total Average Previous Courses Taken	.11	-.09	.04	.01	.12	.04	-.08	.02
R^2	.12	.14	.18	.17	.16	.14	.14	.16
Adjusted R^2	.09	.12	.15	.14	.13	.11	.11	.13

Table 3

Correlations of Student Rate of Repeating the Same Teacher with Measures of Teaching Effectiveness by Students for Mathematics Courses, 1978-1982
N=746

Measures of Teaching Effectiveness	Rate of Students Repeating Same Teacher per Course	
	Course	Total Average Repeats All Courses
1) Well-prepared	.15	.15
2) Stimulated student thinking	.21	.13
3) Helpful to students	.18	.10
4) Explained course objectives	.11	.09
5) Grades fair/impartial	.11	.09
6) Explains difficult material	.12	.06
7) Students learn great deal	.14	.05
Total of 7 Items	.19	.10

References

Abrami, P.C., Dickens, W.J., Perry, R.P., & Leventhal, L. Do teacher standards for assigning grades affect student evaluations of instruction? Journal of Educational Psychology, 1980, 72, (1), 107-118.

Abrami, P.C., Perry, R.P., & Levanthan, L. The relationship between student personality characteristics, teacher ratings, and student achievement. Journal of Educational Psychology, 1982, 74, (1), 11-125.

Braskamp, L.A., Ory, J.C., & Pieper, D.M. Student written comments: Dimensions of instructional quality. Journal of Educational Psychology, 1981, 73, (1), 65-70.

Dowell, D.A., & Neal, J.A. A selective review of the validity of student ratings of teaching. Journal of Higher Education, 1982, 53, (1), 51-62.

DuCette, J., & Kenny, J. Do grading standards affect student evaluations of teaching? some new evidence on an old question. Journal of Educational Psychology, 1982, 72, (3), 308-314.

Hills, J.R. On the use of student ratings of faculty in determination of pay, promotion and tenure. Research in Higher Education, 1974, 2, (4), 317-324.

Hofman, J.E., & Kremer, L Attitudes toward higher education and course evaluation. Journal of Educational Psychology, 1980, 72, (5), 610-617.

Howard, G.S., & Maxwell, S.E. Correlation between student satisfaction and grades: A case of mistaken causation? Journal of Educational Psychology, 1980, 72, (6), 810-820.

Howard, G.S., & Maxwell, S.E. Do grades contaminate student evaluations of instruction? Research in Higher Education, 1982, 16, (2), 175-188.

Marsh, H.W. The influence of student, course, and instructor characteristics in evaluations of university teaching. American Educational Research Journal, 1980, 17, (1), 219-237.

Marsh, H.W. Validity of students' evaluations of college teaching: A multitrait, multimethod analysis. Journal of Educational Psychology, 1982, 74, (2), 264-279.

Overall, J.U., & Marsh, H.W. Relative influence of course level, course type, and instructor on students' evaluation of instruction. Paper presented at the meeting of the American Educational Research Association, Boston, April 1980. (Eric Document Reproduction Service No. ED 187 176)

Yunker, J.A. Validity research on student evaluations of teaching effectiveness: Individual student observations versus class mean observations. Research in Higher Education, 1983, 19, (3), 363-379.